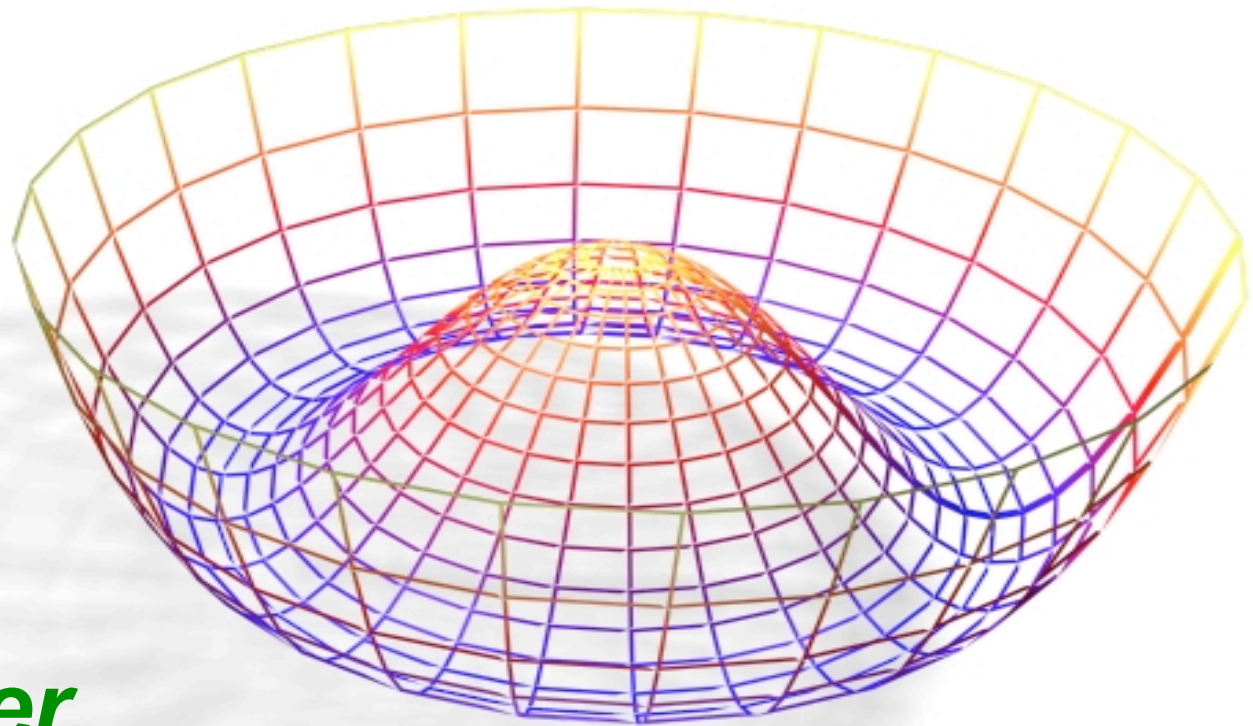


Status of the Atlas analysis tests with PROOF



Kyle Cranmer

Role of PROOF in Analysis Model

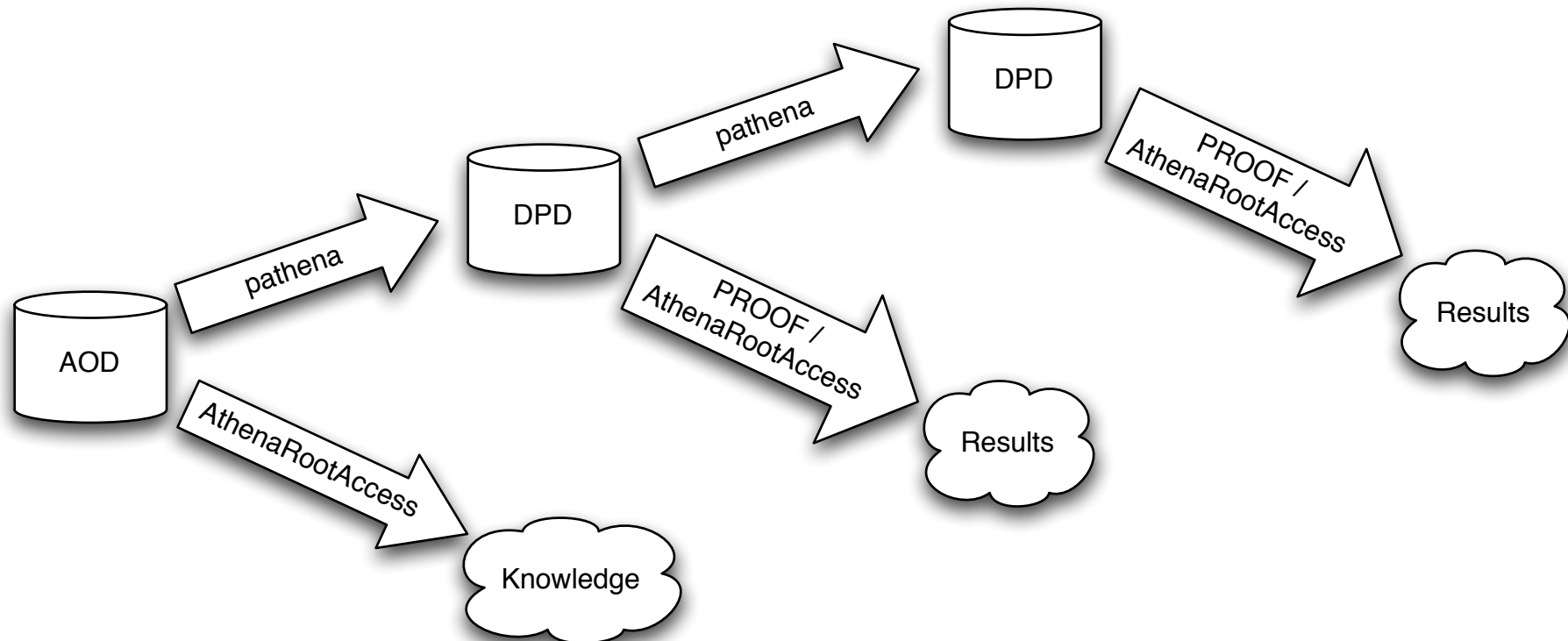
AOD-based analysis with EventView and other tools fairly well-developed.

- pathena is an excellent distributed analysis tool for this.

Output of 1-stage analysis is DPD

- DPD can be flat AANT or POOL-based DPD together with AthenaRootAccess

PROOF an ideal tool for ROOT-based analysis stage (?)



A timeline of recent work and future goals

- Sept 13 & 14 is Atlas Analysis Forum

The ProofTestBed Wiki page

Some initial results on the testbed

Some usability improvements

Milestones for the next month

April 20: initial expression of interest in PROOF for DPD analysis.

April 23: meet with Sergey to setup PROOF with ROOT 5.10

June 18: first successful run in 5.14, send email to broader list

June 19: phone meeting with BNL + Kaushik

39 sec (PROOF) vs.
7 min (local)!

June 21: Torre's talk in Bloomington

June 25: Sergey's talk at SW Week

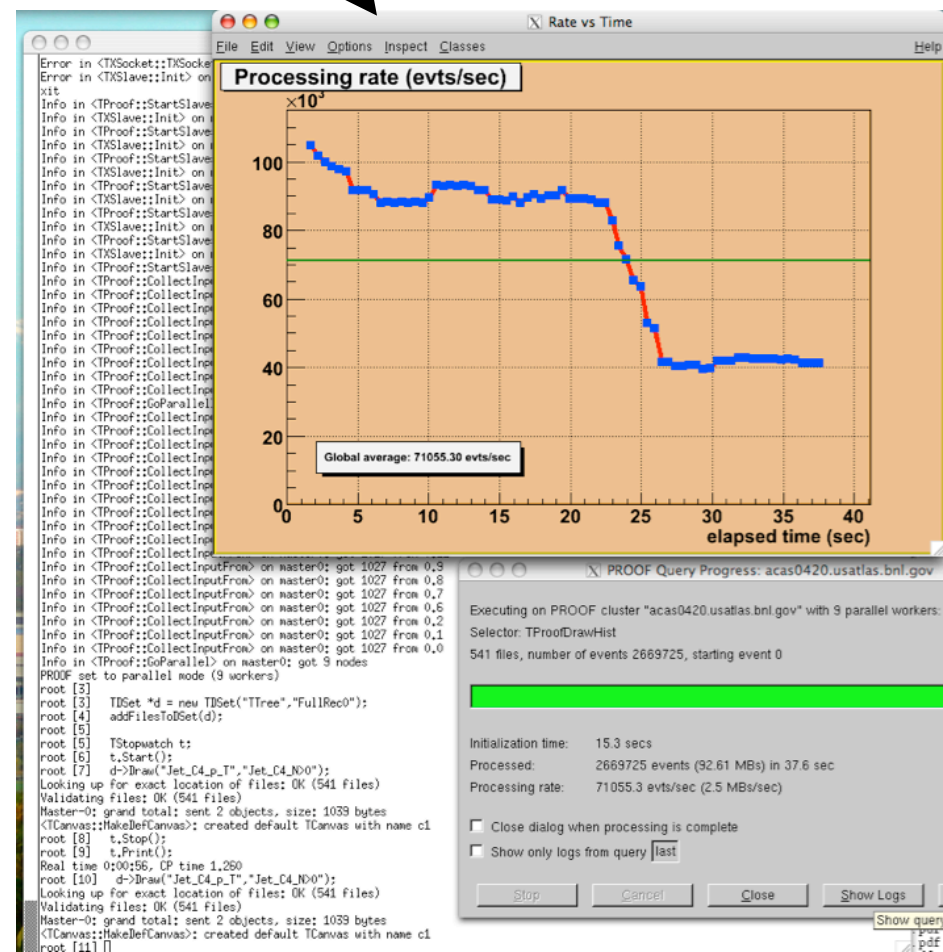
June 30: meet with Michael Ernst, Jim Shank, and Sergey to discuss planning

July 11: Steve Gowdy's talk at US Atlas Overview

July 13: create ProofTestBed wiki page, invite a few more users

July 15–Aug 1: testing, feedback, usability improvements, etc.

Sept 13–14: Input to Analysis Forum Meeting



We created a wiki page to collect documentation, results, feedback, and milestones

<http://www.usatlas.bnl.gov/twiki/bin/view/AtlasSoftware/ProofTestBed>

The Documentation has been followed by a few users without problems.

Initial feedback:

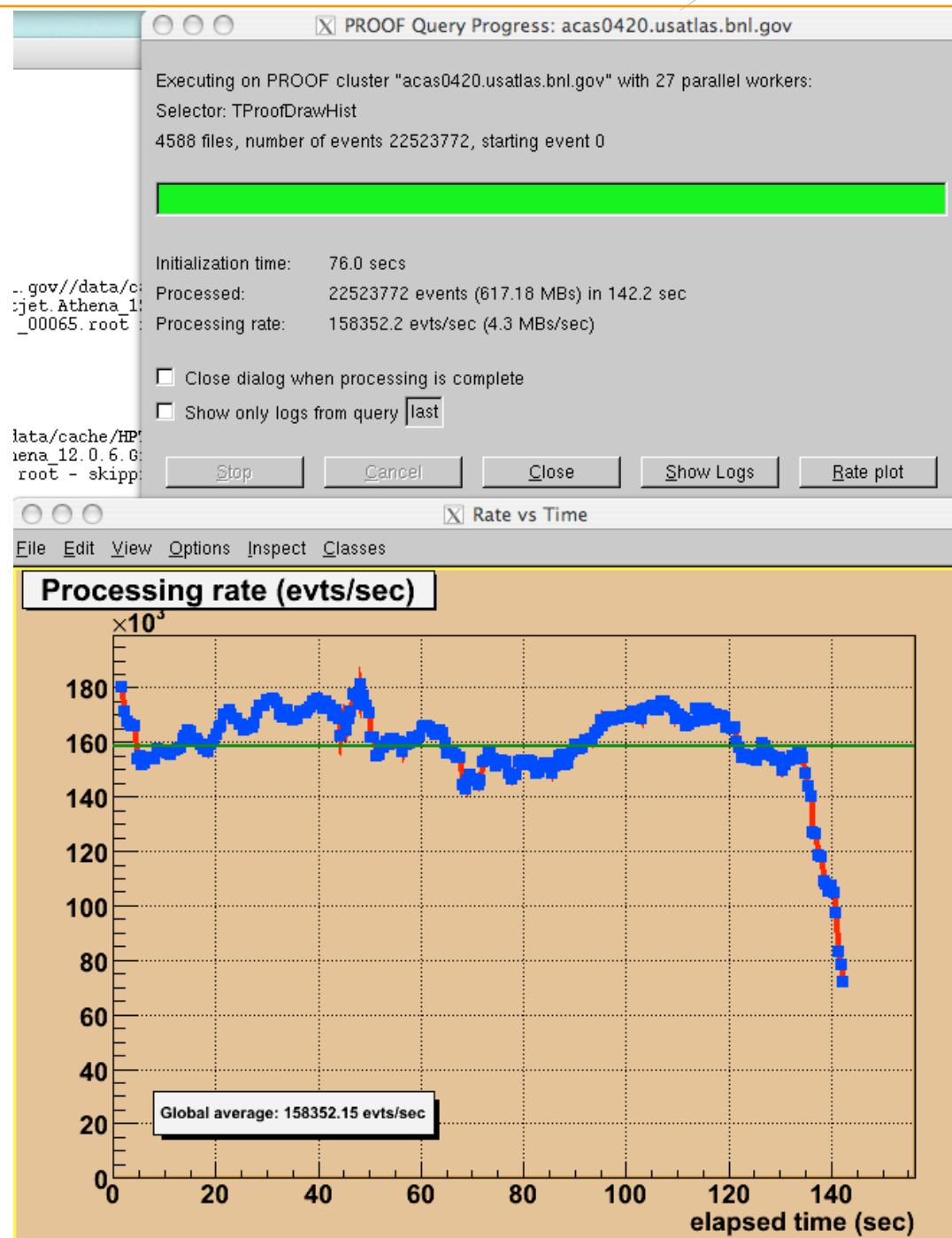
From: kblack@fas.harvard.edu
Subject: Re: Proof at BNL
Date: July 16, 2007 5:08:07 PM EDT
To: cranmer@cern.ch
Cc: panitkin@bnl.gov

wow that is wicked fast. Do you have instructions on what I need to do to get code written from TSelector to run properly. I just took your code and stuck in my macro but it didn't seem to process any events for some reason...

Kevin

The results are from running on 22 million events (ntuples produced with HighPtView).

- ▶ For a simple plot reading only one variable, we sustained 160 kevts/sec across the cluster!
- ▶ When using a TSelector and reading all the data in the tree, the rate drops to 2.6 kevts/sec.
- ▶ Real use-cases should be somewhere in this range.



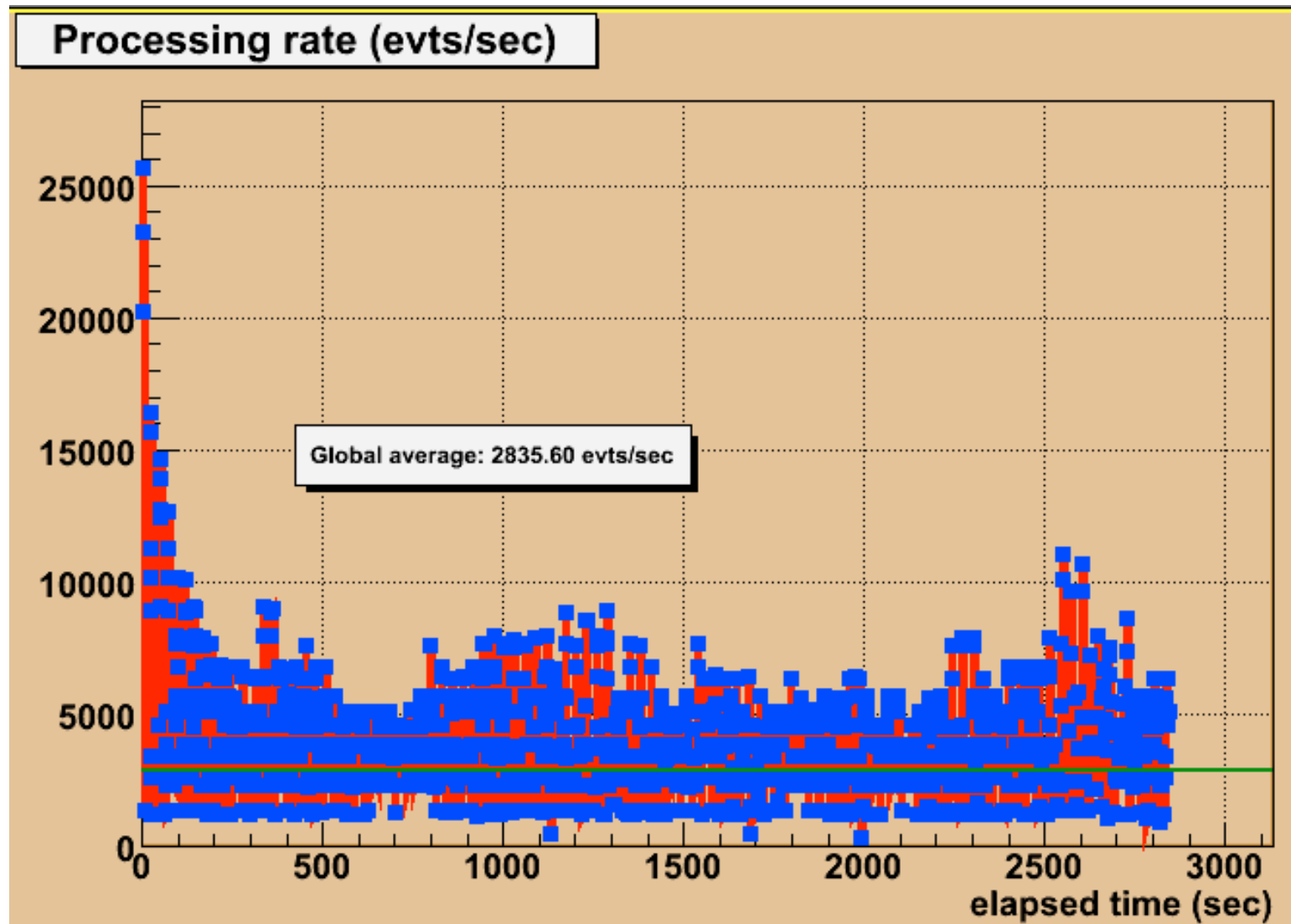
Complicated analysis code must be written in terms of a TSelector to take advantage of PROOF.

- ▶ Fabien first succeeded in getting an example to work on the testbed.
- ▶ Instructions for how to get a TSelector-based analysis are on the wiki page

http://www.usatlas.bnl.gov/twiki/bin/view/AtlasSoftware/ProofTestBed#Example_with_TSelector

- ▶ See a big drop (x50) in processing rate (from 160kevt/s to 3 kevt/s) because the default TSelector reads all data
 - can specify which branches to read
 - see an improvement in I/O by a factor x3

In addition to lower event rate, seems more chaotic

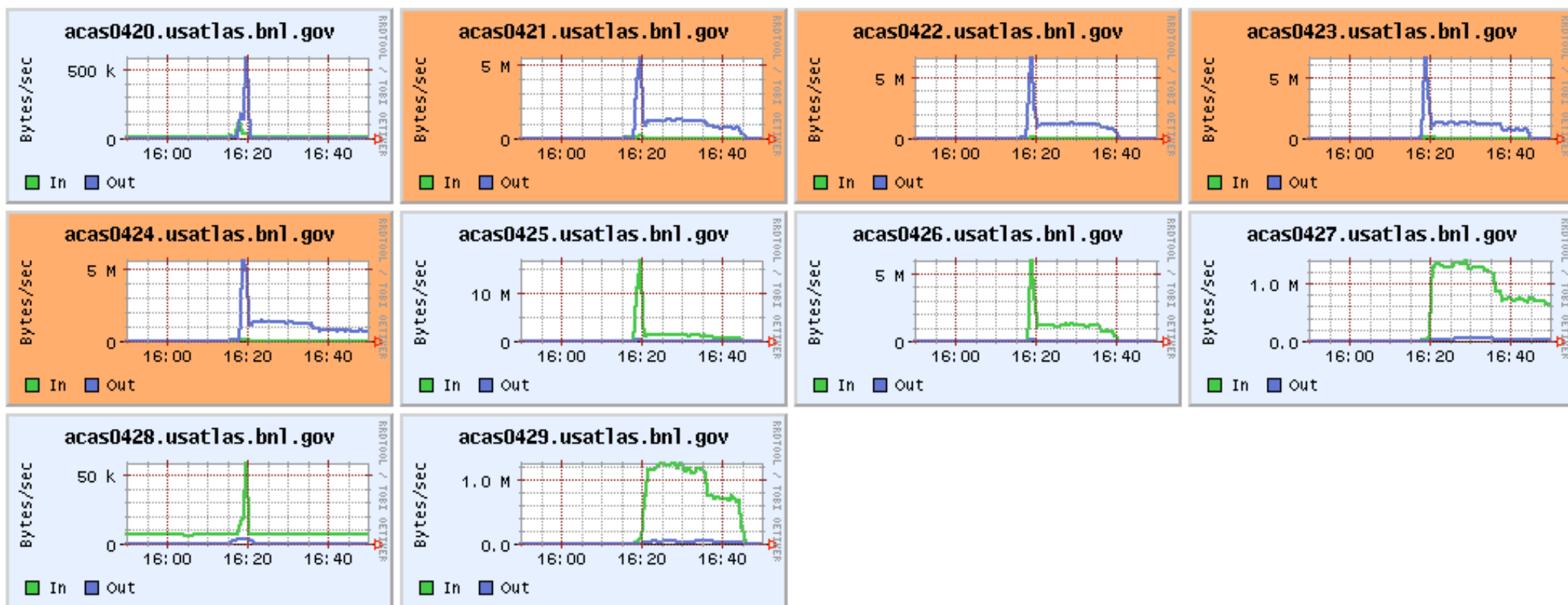


Load-Balance Issues

Initial testbed was populated with files (evenly distributed)

Test bed was expanded (files not re-distributed)

When running on larger testbed see un-even use of resources



We need a convenient way for users to access the files in a dataset (usually produced with pathena)

- Currently we just copy them by hand
- no way to query the files in xrootd

Initial way to add files to TDataSet was a hand-made macro with a long list of files --- no good!

Tadashi, Shuwei, and I created a nicer interface to populate a TDataSet with the files that are in a DQ2 Dataset

http://www.usatlas.bnl.gov/twiki/bin/view/AtlasSoftware/ProofTestBed#Using_DQ2IF_to_add_files_in_a_DQ

```
TDataSet *d = new TDataSet("TTree","FullRec0");

TPython::LoadMacro("DQ2IF.py");
DQ2IF dq2;
TPython::Bind(d,"myTDataSet");
dq2.GetFiles("user.TARRADEFabien.trig1_misal1_mc12.008331.PythiaVBFH120tautaulh.", \
    "myTDataSet" ,"root://acas0420.usatlas.bnl.gov//data/cache/HPTV/","AANT0");
```

Eventual goal is that pathena output directed automatically (either via direct output or via a subscription) to an xrootd-based storage element.

Tadashi to address this in the next talk.

http://www.usatlas.bnl.gov/twiki/bin/view/AtlasSoftware/ProofTestBed.html#Using_PyROOT_in_PROOF

Hi Wim,

Thanks for the info. I'll update the page following your suggestion.
It would be nice if TSelector_PyProxy is officially provided.

Tadashi

From: WLavrijsen@lbl.gov
Subject: Re: Fwd: using PyROOT in PROOF
Date: Fri, 27 Jul 2007 15:59:08 -0700 (PDT)

Kyle,

Tadashi implemented a wrapper for TSelector so PROOF can use PyROOT based analysis code. Can you take a look and comment on it. See below:

that'd work, but I'd do something more like the examples attached (which'd need writing out for most of the functions (and/or generated if I ever get around to do that) and then included with the release or something).

Later,
Wim

Moved from 5.10 to 5.14 because of major improvements in functionality

But, 5.14 has a few problems:

▸ DQ2IF

Have you tried with root-5.15 or root-5.16? I tried 3 versions of root: 5.14.00, 5.15.06 and 5.16.00, and found that python class works only for 5.14.00, but makes root crash for the other two when a python class is instantiated. <https://savannah.cern.ch/bugs/index.php?28179>

▸ PyROOT

Caveat

- The following doesn't work with v.5.14.00g/v5.16.00 because `TSelector::GetSelector("PyTest")` returns NULL due to `G__ClassInfo("PyTest").IsBase("TSelector")==0` and `G__ClassInfo("PyTest").New()==NULL`

```
$ root.exe
root [0] TPython::LoadMacro("PyTest.py")
root [1] ....
root [N] TDataSet *d = ...
root [M] d->Process("PyTest")
```

- PyTest.py would be uploaded to the cluster using `TProof::Load()` with v5.16.00
- `LD_LIBRARY_PATH` and `PYTHONPATH` could be set on the server side

Need to discuss how/when to migrate ROOT versions.

To Do For Analysis Model

- simple macro to translate DQ2 dataset output to add files to TDSet -- done
- Access multiple trees, inheritance of selectors
- Tadashi's example to move pathena output directly to xrd storageelement
 - One can specify destination SE in pathena using `--destSE`. Eg. `pathena --destSE SLACXRD`. Then output will be copied to SLAC xrootd SE via subscription.
- make and distribute .par files
 - setup LD_LIBRARY_PATH to load libraries in a given release
 - Shuwei found this: `gProof->AddDynamicPath(const char *libpath)`
- move 13.0.10 AOD onto xrd server
- demonstraite AthenaROOTAccess?

To Do For Test Bed

- [Xrootd_farm_plans.pdf](#): More details in thsi talk
- network optimization
- repeat scalability tests with network
- redundant proof master nodes & redirector
- possible to get data from dcache?
- file management
- system management & security
- conduit / ssh tunnel

Milestones

Milestones for the Sept 13 & 14 Analysis Forum Meeting

- July 20: Moderately complicated TSelector
- July 20: copy some rel 13 AOD to xrootd servers
- July 27: Some benchmark results with HPTV ntuples + TSelector -- done
- July 27: Example with PAR file
- Aug. 3: setting up LDLIBRARYPATH on slaves, attempt AthenaRootAccess?
- Aug 10: pathena setting up DQ2 Subscription to xrootd as a storage element
- Aug 17: Frank & Kyle demonstrating AOD thinning in rel 13.0.20?
- Aug 17: Fabien & Kyle exercising real-world analysis examples for CSC Higgs To Tau Tau note
- Aug 24: new benchmark results for POOL-based DPD access
- Aug 31-Sept 13: expand testbed, invite more users. iterate on above, producing presentations for meeting